

MPhil in Advanced Computer Science

Introduction to Natural Language Processing

Leader: Ted Briscoe, Stephen Clark, Ann Copestake, Simone Teufel
Timing: Michaelmas term
Prerequisites: None
Structure: 16 Lectures

AIMS

This module aims to provide a brief introduction to linguistics for computer scientists and then goes on to cover some of the core tasks in natural language processing (NLP), with the emphasis on statistical techniques suitable for the extraction of meaning from large bodies of text. Finally we will consider some applications and evaluate how well they work given current techniques.

SYLLABUS

1. Linguistics for NLP – morphology, syntax, semantics and pragmatics with illustrated application mostly to English (6ls, Prof Ted Briscoe)
2. Finite-State / Markovian Techniques – lemmatisation, part-of-speech (PoS) tagging, phrase chunking and named entity recognition (NER) (4Ls, Dr Stephen Clark)
3. Parsing – grammars, treebanks, representations and evaluation, statistical parse ranking (3Ls, Dr Stephen Clark)
4. Interpretation – compositional semantics and entailment, pragmatic inference (2Ls, Dr Ann Copestake)
5. Applications – information extraction, question answering, text classification and their evaluation (1L, Dr Simone Teufel)

OBJECTIVES

On completion of this module students should:

- understand the basic properties of human languages and be familiar with descriptive and theoretical frameworks for handling these properties.
- understand the design of tools for basic NLP tasks such as tagging and (partial) parsing and be able to apply them to text and evaluate their performance.
- understand some of the basic principles of the representation of linguistic meaning and interpretative inference.
- understand how some common applications can be constructed using NLP tools and how their performance can be evaluated.

PRACTICAL WORK

- Practical Week 6: (6 Hours) PoS tag, chunk, and/or perform NER on a designated text with one or more provided tools. Evaluate the performance of the tools quantitatively and qualitatively.
- Practical Week 8: (6 Hours) Parse a designated set of sentences with one or more provided tools, to yield representations of their grammatical relations, phrase structure and/or logical forms. Evaluate the performance of the tools quantitatively and qualitatively.

ASSESSMENT

- There will be four ticked, short, take-home assignments on linguistic analysis during weeks 1-3. Each assignment is worth 5% of the final mark.
- An assessed practical report based on the practicals described above. The practical report will consist of a description of the work done of not more than 5000 words. It will contribute 80% of the final mark.
- The ticked assignments and practical reports will be set and marked by Prof Briscoe and Dr Clark.

RECOMMENDED READING

Jurafsky, D. & Martin, J. *Speech and Language Processing*, Prentice-Hall, 2nd edition, 2008

Last updated: October 2009